CAPÍTULO 2.2

RECOGIDA Y TABULACIÓN DE LA INFORMACIÓN

I.— CONCEPTOS DE TABULACIÓN Y FRECUENCIA

Se entiende por *tabulación* la operación de resumir de manera ordenada un conjunto de datos numéricos en una tabla.

Por otra parte, la *frecuencia* se define como el número de veces que tiene lugar la observación de un determinado fenómeno. Por consiguiente, cuando mencionamos el concepto *tabla de frecuencias*, nos referimos a una tabla en la que se expresa de manera ordenada un conjunto de observaciones con el número de veces que éstas se repiten.

II.— RECOGIDA, ORDENACIÓN Y PRESENTACIÓN DE LA INFORMACIÓN MEDIANTE TABLAS DE FRECUENCIAS

El uso de las tablas de frecuencias es fundamental para poder resumir la información por medio de la cual se estudiará un determinado fenómeno.

Se pueden construir tablas de frecuencias para cualquier tipo de carácter (cualitativo o cuantitativo); además, estas tablas podrán tener una o varias entradas (formas de entrar en la tabla, ya sea mediante una, dos o más variables).

Los elementos que componen una tabla de frecuencias son los siguientes (fig. 2.2):

a) Columnas: están formadas por una serie de datos numéricos o alfabéticos colocados de arriba abajo. La primera columna hace referencia a los valores de la variable que se está tabulando. Cuando haya más de una variable, la primera columna también hará referencia a los valores de una variable. Generalmente, en caso de tablas de entrada única (una sola variable), en las cabeceras de las

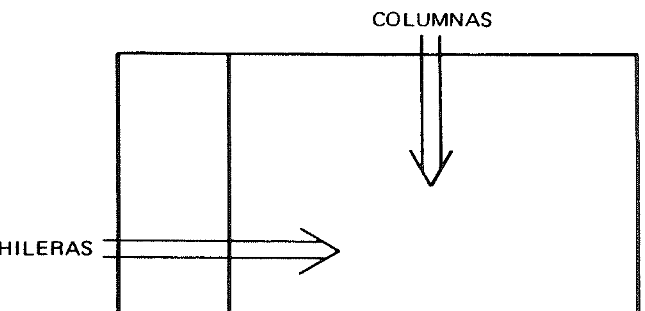


FIG. 2.2.— *Tabla de frecuencias.*

columnas se indica mediante palabras o símbolos la información que se expone en el resto de la columna.

b) Líneas o hileras: están formadas por una serie de datos numéricos o alfabéticos colocados de izquierda a derecha. La primera hilera o línea corresponde a los encabezamientos de cada columna. En el caso de una tabla de entrada única, el primer dato de cada hilera o línea coincide con los valores de la variable que se está tabulando.

Al trabajar con una serie estadística de muchos datos, se recomienda agrupar dichos valores en *intervalos*; esto es aplicable a las variables cuantitativas continuas y, ocasionalmente, a las cuantitativas discretas. El caso más sencillo es la agrupación de los valores de una variable por intervalos en una tabla de entrada única.

Por ejemplo, en un estudio sobre la variable tasa de glucemia puede interesar agrupar dichos valores en unos intervalos definidos con anterioridad y que podrían ser los siguientes: 70-79; 80-89; 90-99; 100-109; 110-119; 120-129 y 130-139. Estas «agrupaciones» se colocarán en la primera columna, y cada una de ellas será el primer dato de cada hilera (tabla 2.2).

TABLA 2.2. -- **Tabla de frecuencias**

INTERVALOS DE LA VARIABLE (TASA DE GLUCEMIA)	
70-79	
80-89	
90-99	
100-109	
110-119	
120-129	
130-139	

TABLA 2.6. — **Tabla de doble entrada**

CATEGORÍAS DE LA VARIABLE I / INTERVALOS DE CLASE Y DE LA VARIABLE Z	A	B	C	D	E
—					
—					
—					
—					
—					
—					
—					
—					

V.— TRANSFORMACIONES BÁSICAS DE LAS FRECUENCIAS

Las frecuencias también se pueden expresar de otras maneras, lo que implica una transformación de la información inicial. Así, podemos hablar de:

1. **Cociente o razón** (*ratio* en inglés).— Es el término genérico para expresar una determinada relación. En la práctica se emplea cuando las magnitudes son independientes:

$$R = \frac{a}{b} \qquad (2.6)$$

Por ejemplo, se puede calcular la razón o cociente entre los varones y las mujeres (razón de masculinidad) de una población determinada:

$$\text{Razón} = \frac{V}{M} = \frac{4.500}{5.500} = 0{,}90$$

O también la razón inversa, entre las mujeres y los varones (razón de feminidad).

2. **Proporción** (*Proportion* en inglés).— Es una razón particular en la que el numerador del cociente es una magnitud incluida en el denominador:

$$P = \frac{a}{a+b} \qquad (2.7)$$

Por ejemplo, se puede calcular la proporción de varones en una determinada población:

$$P = \frac{\text{Varones}}{\text{Población general}} = \frac{4.500}{9.500} = 0{,}47$$

La proporción es sinónimo de frecuencia relativa, tanto por uno o probabilidad.

3. **Porcentaje** (*Percentage* en inglés). — Es una proporción multiplicada por 100:

$$\% = \frac{a}{a+b} \times 100 \qquad (2.8)$$

En el caso del ejemplo anterior, sería: $0{,}47 \times 100 = 47\ \%$.

4. **Tasa** (*rate* en inglés).— Es una proporción particular en la que se quiere resaltar la frecuencia relativa de aparición de un hecho sanitario o la noción de un riesgo de un grupo determinado de la población general en un tiempo dado.

Siempre va multiplicada por un factor *K*, que es múltiplo de 10 (1.000, 10.000, etcétera):

$$T = \frac{a}{a+b} \times K \qquad (2.9)$$

Por ejemplo, se puede calcular la tasa de mortalidad en una población determinada durante un año:

$$\text{Tasa de mortalidad} = \frac{\text{Número de fallecimientos en un año}}{\text{Población general durante ese año}} \times 1.000 =$$

$$= \frac{950}{9.500} \times 1.000 = 10 \text{ por mil}$$

VI.— CRITERIOS BÁSICOS EN LA PRESENTACIÓN DE LA INFORMACIÓN MEDIANTE TABLAS DE FRECUENCIAS

Además de considerar unas normas básicas en la construcción de las tablas de frecuencias, deben tenerse en cuenta otros criterios en la presentación de las mismas; a saber:

1.— Las tablas de frecuencias deben estar claramente rotuladas y bien encabezadas.

2.— Los valores totales deben indicarse correctamente.

3.— En el caso de que se realicen transformaciones básicas de las frecuencias, debe indicarse claramente la base o denominador utilizado para el cálculo.

4.— Si la tabla refiere variables cuantitativas con unidades de medida, éstas deben indicarse de forma visible.

Los elementos a tener en cuenta en una tabla de frecuencias de entrada única y con los datos agrupados por intervalos (también denominados intervalos de clase) son:

1. **Número de intervalos.**— No existe ninguna norma estricta a este respecto, de manera que, en teoría, se puede trabajar sobre un número indeterminado de intervalos. En la práctica, el investigador decide *a priori* el número de intervalos que desea tener y, a partir de esa opción, construye la tabla de frecuencias. Se ha dicho en alguna ocasión que el número de intervalos óptimo sería igual a la raíz cuadrada del número de observaciones realizadas; sea como fuere, esta regla es más bien orientativa y no debe utilizarse de manera estricta*. En la práctica, no es aconsejable establecer más de 10 intervalos. Una vez decidido el número de intervalos de clase que formarán la tabla de frecuencias, se colocan en la primera columna.

En el ejemplo anterior, si se ha determinado la glucemia en 50 individuos, el número teórico de intervalos sería la raíz cuadrada de 50, es decir 7,07; en la práctica, se concretarán en siete u ocho intervalos.

2. **Límites de los intervalos de clase.**— Son los extremos de cada uno de los grupos formados con los valores de la variable; obviamente, el *límite superior* y el *límite inferior* se corresponden con los valores máximo y mínimo de cada intervalo. Además, es necesario distinguir entre *límites reales* y *límites exactos*. Denominamos límites reales a los valores extremos que coinciden con valores observados en alguna situación; entre el límite real superior de un intervalo y el límite real inferior del siguiente hay una diferencia de una unidad. En cambio, denominamos límites exactos a los valores «imaginarios» que, por efectos del redondeo, delimitan con mayor precisión el inicio y final de cada intervalo. Así, el límite exacto inferior queda media unidad por debajo del límite real inferior, y el límite exacto superior media unidad por encima del límite real superior. El límite exacto superior de un intervalo coincidirá con el límite exacto inferior del siguiente. De esta manera, siempre podrá colocarse una determinada medición en algún intervalo de clase.

Por más preciso que sea el instrumento de medida, siempre deben especificarse los límites exactos de los intervalos, con lo que se evitarán posibles errores en la tabulación. Se simbolizan por L_s (límite superior) y L_i (límite inferior); además, puede concretarse si son exactos o reales.

3. **Amplitud de un intervalo.**— Es la distancia entre el límite exacto inferior y el límite exacto superior de un intervalo. También se denomina rango o recorrido del intervalo. Según este concepto, podemos distinguir dos tipos de tablas de frecuencias: *a*) con intervalos de igual amplitud, y *b*) con intervalos de distinta amplitud. Se simboliza por *I*.

4. **Punto medio del intervalo.**— Es el valor promedio, o centro de gravedad, de todos los valores que componen un determinado intervalo de clase. En la práctica,

*Otra regla para calcular el número de intervalos es la siguiente:

$$\text{n.º de intervalos} = 1 + \frac{\log N}{0,3}$$

donde N es el tamaño de la muestra.

se calcula por la semisuma de los límites superior e inferior (ya sean los exactos, ya los reales). Se simboliza por las siglas PMI.

Continuando con el ejemplo de la tasa de glucemia en una muestra de 50 individuos, se puede confeccionar la siguiente tabla de frecuencias de entrada única (tabla *2*.3), en la que aparecen los términos citados hasta ahora.

TABLA *2*.3. **Tabla de frecuencias**

INTERVALOS DE LA VARIABLE (TASA DE GLUCEMIA)	PMI	I
(69,5) 70-79 (79,5)	74,5	10
(79,5) 80-89 (89,5)	84,5	10
(89,5) 90-99 (99,5)	94,5	10
(99,5) 100-109 (109,5)	104,5	10
(109,5) 110-119 (119,5)	114,5	10
(119,5) 120-129 (129,5)	124,5	10
(129,5) 130-139 (139,5)	134,5	10

Entre paréntesis se indican los límites exactos de cada intervalo.

III.— TIPOS DE FRECUENCIAS

Existen diferentes criterios para clasificar las frecuencias. En primer lugar, cabe citar la «relatividad» de las frecuencias como criterio clasificatorio; así, hablaremos de *frecuencias absolutas* cuando nos refiramos al número de veces que se repite el valor de una variable, y de *frecuencias relativas* cuando expresemos las frecuencias absolutas con respecto al total de individuos estudiados o analizados.

Un segundo criterio clasificatorio es el de «acumulación» de las frecuencias; en este caso, hablaremos de *frecuencias* simples cuando expresemos solamente la ocurrencia de un determinado valor de la variable estudiada, y de *frecuencias acumuladas* cuando nos interese dar la noción de «sumatorio» o «sumación» hasta un determinado valor de la variable. Obviamente, estos valores han de haber sido ordenados previamente.

Combinando estos dos criterios, obtenemos cuatro tipos de frecuencias (tabla *2*.4):

a) *Frecuencias absolutas simples* (n_i o FAS). Son aquellas que expresan el número de veces que ocurre un determinado valor de la variable, y sólo dicho valor.

b) *Frecuencias relativas simples* (f_i o FRS). Se calculan dividiendo las frecuencias absolutas simples por el número total de individuos estudiados.

$$f_i = \frac{n_i}{N} \qquad (2.1)$$

Proporcionan una idea «relativa» del número de veces que ocurre un determinado valor de la variable, y sólo dicho valor. Hay que hacer notar que el término frecuencia relativa simple es sinónimo de probabilidad, proporción o tanto por uno.

c) Frecuencias absolutas acumuladas (N_i o FAA). Son aquellas que expresan el «sumatorio» o «sumación» del número de veces que ocurren determinados valores de la variable hasta un valor dado; evidentemente, los valores de la variable deben estar ordenados. Se calculan mediante la siguiente expresión:

$$N_i = \Sigma n_i \qquad (2.2)$$

d) Frecuencias relativas acumuladas (F_i o FRA). Se calculan dividiendo las frecuencias absolutas acumuladas por el número total de individuos observados, de acuerdo con la siguiente expresión:

$$F_i = \frac{N_i}{N} \qquad (2.3)$$

TABLA 2.4. — **Tipos de frecuencias**

FRECUENCIAS	SIMPLES	Absolutas (n_i)
		Relativas ($f_i = \frac{n_i}{N}$)
	ACUMULADAS	Absolutas (N_i)
		Relativas ($F_i = \frac{N_i}{N}$)

Al hablar de frecuencias también hemos de citar los conceptos de porcentaje y percentiles.

Los *porcentajes* (%) se pueden definir como la expresión matemática de las frecuencias relativas simples (o proporciones) multiplicadas por cien. En ocasiones también se denominan *índices*. Se calculan mediante la siguiente fórmula:

$$\% = f_i \times 100 \qquad (2.4)$$

En la práctica, facilitan la comprensión de las frecuencias simples al suprimir los números decimales o fraccionados.

En la tabla 2.5 se ejemplifica la tabla de frecuencias resultante del estudio de la tasa de glucemia en 50 individuos.

TABLA 2.5. — **Tabla de frecuencias de la distribución de la tasa de glucemia en una muestra**

INTERVALOS	PMI	I	n_i	f_i	%	N_i	F_i
70-79	74,5	10	3	0,06	6	3	0,06
80-89	84,5	10	5	0,10	10	8	0,16
90-99	94,5	10	10	0,20	20	18	0,36
100-109	104,5	10	14	0,28	28	32	0,64
110-119	114,5	10	10	0,20	20	42	0,84
120-129	124,5	10	5	0,10	10	47	0,94
130-139	134,5	10	3	0,06	6	50	1
			50	1	100		

Los *percentiles* (P_i) se pueden definir como la expresión matemática de las frecuencias relativas acumuladas (o probabilidades acumuladas) multiplicadas por cien.

Se calculan mediante la siguiente fórmula:

$$P_i = F_i \times 100 \qquad (2.5)$$

IV.— NORMAS PARA LA CONSTRUCCIÓN DE UNA TABLA DE FRECUENCIAS

Para elaborar una tabla de frecuencias que sea clara y no se preste a falsas interpretaciones, deben tenerse en cuenta las siguientes normas:

1. El número de intervalos de clase debe oscilar alrededor de 10. Recordar, como guía orientativa, que dicho número puede ser aproximadamente la raíz cuadrada del total de individuos observados.

2. En el caso de trabajar con intervalos de clase, los límites reales deben coincidir con la precisión de los datos originales.

3. Conviene utilizar intervalos de igual amplitud, ya que facilitan los cálculos posteriores y la construcción de gráficos a partir de dicha tabla.

4. Los intervalos de clase deben ser mutuamente excluyentes, de forma que un determinado valor de la variable pueda ser incluido únicamente en un intervalo.

Ejemplos:	*Intervalo*	*Ejemplos:*	*Intervalo*
a) No es correcto:	70-79	*c)* Es correcto:	70- <80
	79-88		80- <90
	88-97		90- <100
	⋮		⋮
b) Es correcto:	70-79	*d)* Es correcto:	70-79,99
	80-89		80-89,99
	90-99		90-99,99
	⋮		⋮

5. Debe evitarse en lo posible el tratamiento de los datos mediante intervalos abiertos. Un ejemplo ya clásico es la tabla de frecuencias que refleja la estructura de una población según la variable edad; en ella el último intervalo suele dejarse abierto y expresarse como: +75 años; 75 años y más, o ≥75 años.

6. Para realizar cálculos posteriores a partir de los datos correspondientes a la distribución de frecuencias, se puede determinar el punto medio de cada intervalo de clase (PMI). Dicho punto se utilizará como valor representativo del intervalo.

7. Recordar que las tablas de frecuencias pueden ser de entrada única, doble e incluso múltiple, aunque estas últimas son de más difícil interpretación (tabla 2.6).

5. — A menudo las tablas de frecuencias permiten expresar los resultados de una forma más concisa y clara que un texto.

6. — Las tablas de frecuencias excesivamente complejas deben resultar inteligibles.

VII.— NÚMEROS ÍNDICES

Los números índices pueden considerarse como una transformación básica de las frecuencias. Constituyen un tipo especial de las proporciones y, por extensión, de las frecuencias relativas simples. La característica que los distingue de la clasificación realizada en el apartado V es su aplicación a las series cronológicas (véase capítulo *3.8*), de uso generalizado en disciplinas como la economía, la demografía, etc.

Los números índices se calculan a partir del valor de la variable en un momento determinado de la serie, que se toma como base y que generalmente se equipara a 100. Todos los demás valores se calculan como porcentajes de este valor inicial.

El coste de la vida, el nivel de los salarios y el comercio exterior son ejemplos de fenómenos económicos que se pueden medir mediante los números índices. Un ejemplo de interés sanitario viene representado en la tabla *2.7*.

TABLA *2.7.* — **Consumo de cigarrillos per cápita de la población adulta de Cataluña durante el período 1973-1980 (año base: 1973)**.

AÑO	n_i CATALUÑA	n.° índice
1973	2.450 cigar/cápita	100
1974	2.476 cigar/cápita	101,06
1975	2.304 cigar/cápita	94,04
1976	2.434 cigar/cápita	99,35
1977	2.592 cigar/cápita	105,96
1978	2.306 cigar/cápita	94,12
1979	2.576 cigar/cápita	105,14
1980	2.492 cigar/cápita	101,71

Fuente: «El tabaquismo en Cataluña». Departamento de Sanidad y Seguridad Social. Generalitat de Catalunya. Barcelona, 1983, p. 251 (modificada.)

En este caso se puede afirmar que en el año 1974 se consumió en Cataluña un 1,06 % más de cigarrillos per cápita que en 1973.

CAPÍTULO 2.4

REPRESENTACIÓN GRÁFICA DE LA INFORMACIÓN

I.— INTRODUCCIÓN

Cuando el investigador pretende *describir* de forma clara y resumida la información recogida sobre la variable a estudio en una determinada muestra o grupo, recurre a la «imagen». El aserto «una imagen vale más que mil palabras» se puede aplicar al ámbito de la estadística descriptiva diciendo que «un gráfico bien elaborado vale más que mil tablas de frecuencias». Cada vez es más habitual el uso de gráficos o imágenes para presentar la información obtenida. De todas maneras, se debe ser prudente al confeccionar o interpretar gráficos, puesto que una misma información se puede representar de formas muy diversas, y no todas ellas son pertinentes, correctas o válidas.

Los publicistas saben muy bien cómo presentar un determinado producto para que cause impacto entre la población a la cual va dirigido. En la descripción de datos estadísticos también puede ocurrir que una determinada representación gráfica o imagen se haya constituido de forma equívoca para dar una impresión tergiversada de la información. Por supuesto, los motivos de tal conducta son ajenos a la estadística en sí.

Por todo ello, es preciso dar unos criterios y normas mínimas para construir y presentar adecuadamente los gráficos en estadística descriptiva.

Actualmente, la mayoría de gráficos estudiados en este capítulo pueden ser construidos con cualquier ordenador que disponga de *plotter* (trazador de gráficas).

II.— REPRESENTACIÓN GRÁFICA DE LOS ATRIBUTOS Y DE LAS VARIABLES CUANTITATIVAS DISCRETAS

Los atributos y las variables cuantitativas discretas se pueden representar mediante los siguientes tipos de gráficos: *a*) diagrama de barras; *b*) superficies representativas, y *c*) diagrama polar. Las superficies representativas, a su vez, se

pueden clasificar según sean sus formas geométricas o geográficas. Las más usadas son los sectores circulares, los pictogramas y los cartogramas.

A.— DIAGRAMA DE BARRAS

El diagrama de barras es un tipo de representación gráfica apropiado para variables cualitativas y cuantitativas discretas.

Para su construcción se necesitan unos ejes de coordenadas. El eje de abcisas se asigna a los valores de la variable y el de ordenadas a la escala de frecuencias, que pueden ser absolutas o relativas. Sea cual sea la frecuencia considerada, la escala siempre debe inicirse en cero y coincidir con el cero de coordenadas. A continuación, se levantan líneas perpendiculares (barras) sobre el eje de abcisas cuya altura será igual a la frecuencia de cada uno de los valores de la variable.

En la figura *2.5* se muestra un ejemplo de diagrama de barras basado en los datos de la tabla *2.10*.

TABLA *2.10*. — **Tipos de escuelas en el Programa de Educación Sanitaria en las escuelas de Cataluña**

TIPOLOGÍA DE LA ESCUELA	n_i	f_i	%
Rural	19	0,373	37,3
Urbana	24	0,470	47,0
Suburbial	8	0,157	15,7
Total	51	1	100

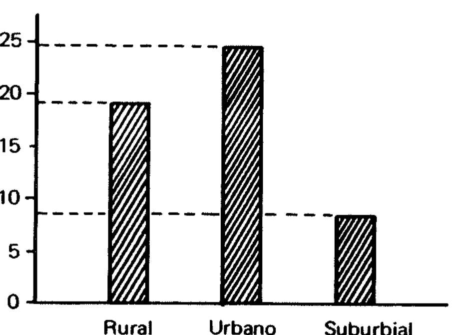


FIG. *2.5*.— *Diagrama de barras de la variable «tipología de la escuela» en el Programa Piloto de Educación Sanitaria de Cataluña.*

B.— SUPERFICIES REPRESENTATIVAS

Las superficies representativas se pueden clasificar en diferentes tipos según las áreas o volúmenes que nos interese utilizar; de todas maneras, en la práctica, las más empleadas son los sectores circulares. En ocasiones pueden resultar

adecuados los cilindros en forma de «tarta» o «pastel», las pirámides, los cubos, los prismas e incluso los pictogramas y cartogramas.

Sectores circulares

Los sectores circulares se construyen representando cada valor de la variable estudiada mediante una porción de círculo. Estos sectores deben ajustarse proporcionalmente a la frecuencia de aparición de los valores de la variable. Para ello, será preciso calcular los grados del ángulo de su sector circular mediante una regla de tres simple o directa. Una vez repartidos los 360° del círculo en los diferentes sectores, uno por cada valor de la variable, se puede trazar la imaen.

El ejemplo utilizado en el caso del diagrama de barras (tabla *2.10*) también se puede representar mediante sectores circulares; veamos su cálculo y construcción:

La regla de tres directa la planteamos del siguiente modo:

51 escuelas . . . . 360°
19 escuelas rurales . . . . $x°$

$$x° = \frac{360° \times 19 \text{ escuelas}}{51 \text{ escuelas}} = 134{,}12°$$

Siguiendo el mismo procedimiento, se obtiene:

Urbana . . . . 169,41°
Suburbial . . . . 56,47°

El gráfico resultante será el de la figura *2.6*.

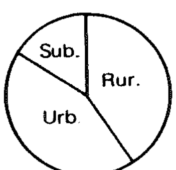


FIG. *2.6*.— *Sectores circulares de la variable «tipología de la escuela» resumida en la tabla 2.10.*

Esta representación gráfica permite colocar en cada sector circular el valor de la variable o su porcentaje. Además, si se dispone de colores, se consigue una imagen de mayor impacto visual.

Pictogramas

Son representaciones gráficas en las que se utilizan las superficies enmarcadas en un dibujo o «pintura». En la práctica, dichos gráficos se concretan en la utilización de «muñecos esquemáticos» dentro de los cuales se señalan los diferentes valores de las variables. A veces, se puede recurrir a una serie de «muñecos» para representar la tendencia de una variable en el tiempo.

El problema que se plantea al intentar representar la evolución cronológica de, por ejemplo, una tendencia demográfica mediante un pictograma es que a

menudo la superficie de cada muñeco no guarda proporción con las frecuencias de un determinado valor de la variable, y ello por la dificultad de calcular el área exacta del dibujo. Esto puede dar una imagen falsa de la información que se pretende presentar.

La figura 2.7 muestra un ejemplo de pictograma a partir de los datos de la tabla 2.11.

TABLA 2.11. — **Evolución de la mortalidad general en Cataluña**

AÑO	TASA DE MORTALIDAD GENERAL EN CATALUÑA (× 1.000 habitantes)
1940	14,0
1945	11,5
1950	11,0
1955	10,2
1960	9,4
1965	8,9
1970	8,4
1975	8,1
1980	7,0

Tabla elaborada con datos del Movimiento Natural de la Población, 1940-1979. Boletín de Estadística y Coyuntura, 1978-180. Anuario Estadístico de España, 1981. Instituto Nacional de Estadística, Madrid.

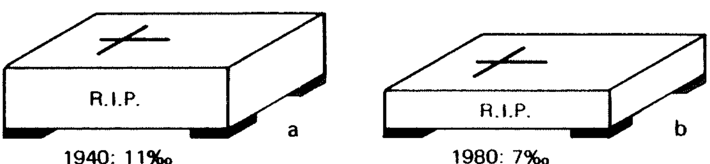


FIG. 2.7.— *Disminución de la mortalidad general en Cataluña durante el período 1940-1980. (Téngase en cuenta que a=b).*

Cartograma

El cartograma se puede definir como aquel gráfico que utiliza una superficie representativa relacionada con mapas geográficos. En la práctica emplea como soporte del impacto visual un mapa en el que se indican las zonas que corresponden a cada uno de los valores de la variable estudiada.

En la figura 2.8 puede verse un ejemplo de cartograma.

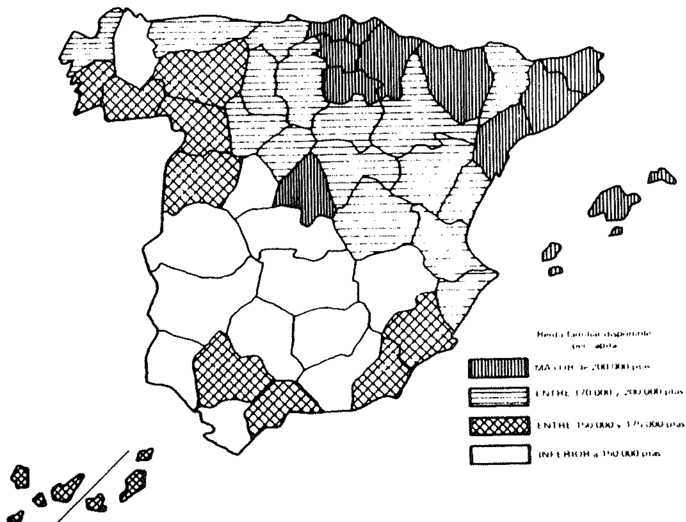


FIG. 2.8.— *Las cuatro capas de la economía española. Fuente: BANCO DE BILBAO: Renta Nacional de España y su distribución provincial, 1977. Servicio de Estudios del Banco de Bilbao. Madrid, 1978, pág. 17.*

C.— DIAGRAMA POLAR

Esta representación gráfica es menos utilizada que las anteriores. Por la forma de presentación —que no por su construcción— es parecido a los sectores circulares.

Para elaborar un diagrama polar se sigue la siguiente sistemática: se toman las frecuencias absolutas o relativas o los porcentajes de cada una de las modalidades del atributo objeto de estudio y, con dichos valores, se fija la longitud de los radios correspondientes, que parten todos ellos de un punto central teórico. De esta forma se «polarizan» todas las modalidades en un mismo punto central, a partir del cual se van trazando gráficamente las rectas correspondientes a los radios. Es recomendable repartir los grados del círculo según el número de modalidades de la variable a estudio.

Ejemplo. Representar mediante un diagrama polar los siguientes datos sobre accidentes de tráfico según los días de la semana:

Días de la semana	%
Lunes	16
Martes	10
Miércoles	9
Jueves	7
Viernes	18
Sábado	21
Domingo	20

El diagrama polar resultante es el mostrado en la figura 2.9.

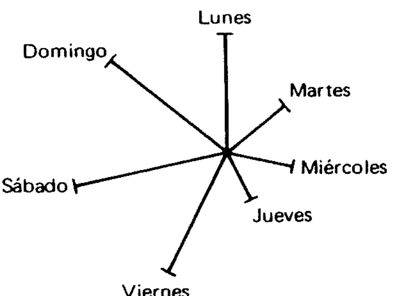


FIG. 2.9. — *Diagrama polar de la variable «accidentes de tráfico según los días de la semana».*

III.— REPRESENTACIÓN GRÁFICA DE LAS VARIABLES CUANTITATIVAS CONTINUAS

Las variables cuantitativas continuas se pueden representar mediante los siguientes tipos de gráficos: *a*) histogramas; *b*) polígonos de frecuencias, y *c*) curvas de frecuencias acumuladas.

A.— HISTOGRAMA

Este tipo de representación gráfica es el más utilizado para las variables cuantitativas continuas dada su relativa sencillez de construcción.

La sistemática que se debe seguir para elaborar un histograma es la siguiente:

1.— Los valores de la variable deben agruparse en intervalos.

2.— Se fijan los límites exactos de cada uno de los intervalos de la variable en el eje de abcisas; no es imprescindible que el cero concuerde con el inicio del eje.

3.— Sobre el eje de ordenadas se construye la escala de frecuencias; aquí sí es necesario que el cero coincida con el inicio del eje.

4.— Sobre el eje de abcisas se levantan tantos rectángulos como intervalos existan. Para que el área sea proporcional a la frecuencia, su altura se calcula de la siguiente manera:

Área = Base × altura

$$\text{Altura} = \frac{\text{Área}}{\text{Base}} = \frac{\text{Frecuencia}}{\text{Amplitud de interalo}} \qquad (2.28)$$

De todo ello se deduce que:

a) Si todos los intervalos en los que se han agrupado los valores de la variable tienen la misma amplitud, las alturas son proporcionales a sus frecuencias, por lo que pueden trasladarse directamente sobre el eje de ordenadas.

b) Si no todos los intervalos en los que se han agrupado los valores de la variable tienen la misma amplitud, las alturas de cada uno de los rectángulos que

forman el histograma, no serán proporcionales a las frecuencias de dichos intervalos. En este caso debe utilizarse obligatoriamente la fórmula 2.28.

c) Si algún intervalo extremo es abierto, resulta imposible construir el histograma al no poder precisar dónde empieza o acaba la distribución de la variable. Una excepción al respecto se encuentra en la construcción de las denominadas «pirámides de población», en las que el intervalo correspondiente al último grupo de edad se delimita convencionalmente en 90 o 95 años.

Veamos ahora un ejemplo de construcción de un histograma. En el capítulo 2.3 ya hicimos mención a la determinación de la variable edad de los 51 directores de las escuelas que habían participado en el Programa Piloto de Educación Sanitaria en Cataluña. Entonces utilizamos dichos valores para elaborar una tabla de frecuencias y calcular las medidas de tendencia central y de dispersión. El histograma de tales datos es el representado en la figura 2.10.

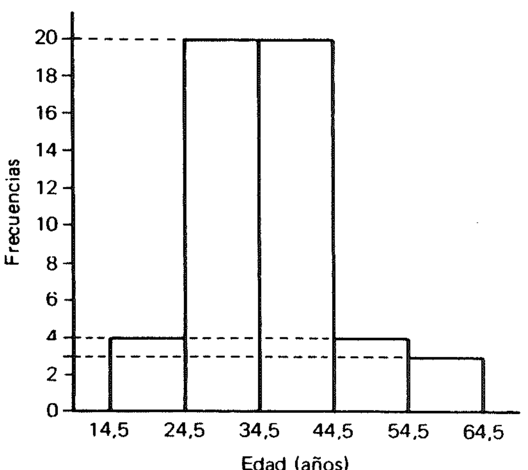


FIG. 2.10. *Histograma de la variable «edad (en años) de los 51 directores de las escuelas del Programa Piloto de Educación Sanitaria en Cataluña».*

B.— POLÍGONO DE FRECUENCIAS

El histograma es la representación gráfica más utilizada en el campo de las variables cuantitativas continuas. De todas formas, a veces se complementan con el denominado «polígono de frecuencias». Este gráfico se construye sobre el histograma; por ello, es condición previa indispensable haber realizado el histograma de la variable continua en cuestión.

Para elaborar un polígono de frecuencias se sigue la siguiente sistemática: en primer lugar, se marca los puntos medios de los intervalos en la base superior de los rectángulos que forman el histograma. A continuación, se unen entre sí estos puntos medios. Por último para cerrar esta «línea quebrada», se unen el punto medio del primer intervalo con el teórico punto medio del intervalo anterior al primero y el punto medio del último intervalo con el teórico punto medio del intervalo posterior a él. Este último paso se podrá realizar siempre y cuando todos

los intervalos tengan la misma amplitud. Esto impide que el polígono de frecuencias se pueda construir cuando los intervalos tienen diferentes amplitudes o existe algún intervalo abierto.

Continuando con el ejemplo utilizado en el histograma, veamos cómo construir el polígono de frecuencias sobre dicha representación gráfica. En la tabla *2.12* se presentan los datos de este ejemplo, con dos intervalos añadidos.

TABLA *2.12*. — **Tabla de frecuencias (modificada) de la variable «edad (en años) de los 51 directores de las escuelas del Programa Piloto de Educación Sanitaria de Cataluña»**

EDAD (en años)			PMI	n_i
(4,5)	5-14	(14,5)	9,5	0
(14,5)	15-24	(24,5)	19,5	4
(24,5)	25-34	(34,5)	29,5	20
(34,5)	35-44	(44,5)	39,5	20
(44,5)	45-54	(54,5)	49,5	4
(54,5)	55-64	(64,5)	59,5	3
(64,5)	65-74	(74,5)	69,5	0
				51

El histograma de estos datos es el mismo que el de la figura *2.10*, en tanto que el polígono de frecuencias es el mostrado en la figura *2.11*.

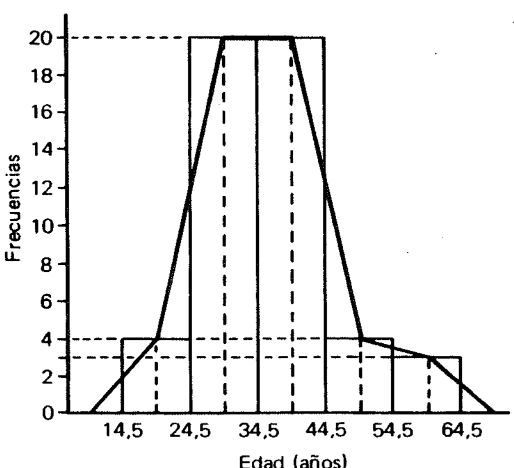


FIG. *2.11*. — *Polígonos de frecuencias de la variable «edad (en años) de los 51 directores de las escuelas del Programa Piloto de Educación Sanitaria de Cataluña».*

C.— CURVA DE FRECUENCIAS ACUMULADAS

Las dos representaciones gráficas anteriores (el histograma y el polígono de frecuencias) sirven para expresar variables continuas cuando se dispone de las frecuencias simples (absolutas, relativas o porcentajes) de cada valor o intervalo. Sin embargo, cuando hay que representar variables continuas con frecuencias acumuladas, se recurre a otros dos tipos de gráficos; a saber, el histograma acumulado y la curva de frecuencias acumuladas. Aunque teóricamente se citan como dos representaciones gráficas distintas, aquí las comentaremos en un mismo apartado, ya que para trazar una curva de frecuencias acumuladas es necesario construir el histograma acumulado.

La sistemática que se sigue para construir ambas gráficas es la siguiente: en primer lugar, se confecciona el histograma acumulado siguiendo las normas citadas anteriormente. La diferencia estriba en el uso de frecuencias acumuladas (absolutas o relativas); esto implica que los rectángulos serán cada vez de mayor altura y que el último será igual a la última frecuencia acumulada; es decir, todos los individuos.

Continuando con el ejemplo anterior, en la tabla *2.13* se expresa el conjunto de frecuencias acumuladas.

TABLA *2.13*. — **Tabla de frecuencias acumuladas de la variable «edad (en años) de los 51 directores de las escuelas del Programa Piloto de Educación Sanitaria de Cataluña»**

EDAD (en años)			PMI	n_i	N_i
(14,5)	15-24	(24,5)	19,5	4	4
(24,5)	25-34	(34,5)	19,5	20	24
(34,5)	35-44	(44,5)	39,5	20	44
(44,5)	45-54	(54,5)	49,5	4	48
(54,5)	55-64	(64,5)	59,5	3	51
				51	

El histograma acumulado correspondiente a dichos datos es el representado en la figura *2.12*.

Una vez realizado el histograma acumulado, ya se puede construir la *curva de frecuencias acumuladas*. Para ello, se deben unir los límites superiores exactos de cada intervalo por el extremo más alto de los rectángulos que conforman el histograma acumulado.

Evidentemente, se obtiene una línea quebrada que va ascendiendo progresivamente y que no llega a cerrarse. Tal es el impacto visual que da este tipo de representación gráfica. Con variables continuas, esta «línea quebrada» se convertirá en una curva a medida que disminuye la amplitud de los intervalos.

Para el ejemplo anterior, la curva de frecuencias acumuladas es la representada en la figura *2.13*.

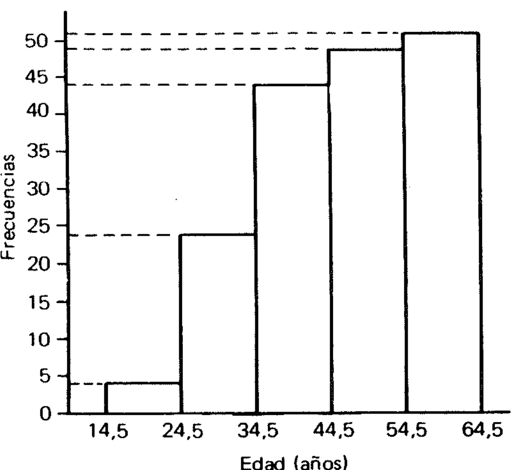


FIG. 2.12.— *Histograma acumulado de la variable «edad (en años) de los 51 directores de las escuelas del Programa Piloto de Educación Sanitaria de Cataluña».*

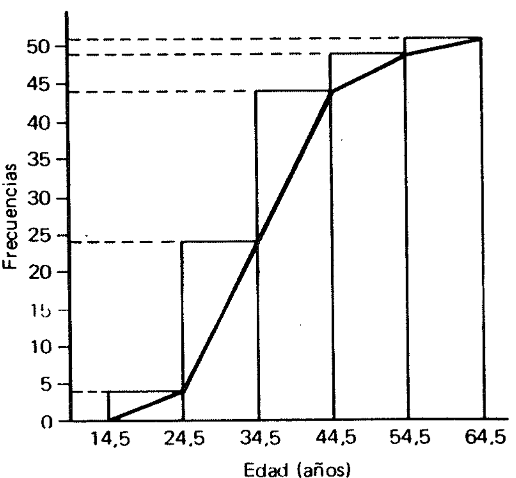


FIG. 2.13.— *Curva de frecuencias acumuladas de la variable «edad (en años) de los 51 directores de las escuelas del Programa Piloto de Educación Sanitaria de Cataluña».*

IV.— OTROS TIPOS DE REPRESENTACIÓN GRÁFICA

Después de comentar los principales tipos de representación gráfica utilizados en el ámbito de la estadística descriptiva, hay que hacer mención de otros gráficos a los que puede recurrirse en determinadas circunstancias: el diagrama de puntos y las gráficas en papel.

A.— DIAGRAMA DE PUNTOS

Este tipo de gráfico no tiene una sistemática de construcción precisa y específica. Su principal ventaja es que puede representar una cantidad importante de información y proporcionar un impacto visual de conjunto sobre el fenómeno a estudio.

Veamos un ejemplo. En la figura 2.14 se representa la evolución de 25 estudiantes de medicina que fumaban durante los 6 años de carrera, haciendo referencia a los que dejaron de fumar (cuadrados) y a los que siguieron con el hábito (círculos). En este caso, los 25 estudiantes no perdieron ningún curso y eran fumadores antes de iniciar la carrera.

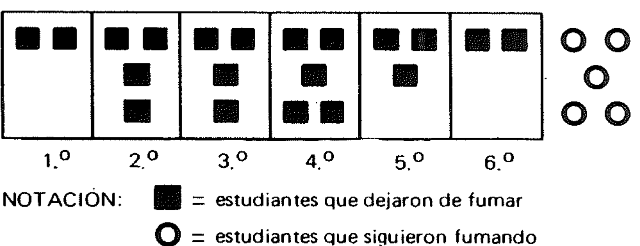


FIG. 2.14.— *Evolución del hábito tabáquico en 25 estudiantes de medicina durante los 6 años de la carrera.*

B.— GRÁFICAS EN PAPEL

Con el término «gráficas en papel» se hace referencia a la representación de la información estadística recogida en un papel previamente «marcado» (cuadriculado, milimetrado, con ciclos logarítmicos...). Estos tipos de gráficos difieren de todos los mencionados hasta el momento por el trazado previo del papel. Por lo general, se representan simultáneamente más de una variable, siempre cuantitativas. Existen tres tipos de gráficas en papel: el gráfico lineal, el gráfico semilogarítmico y el gráfico logarítmico.

Gráfico lineal

Este tipo de gráfico se construye sobre dos ejes que ya suelen estar marcados en el papel. En realidad, son los ejes de coordenadas, con un punto 0 origen de ambos ejes: el horizontal o eje de abcisas (X), y el vertical o eje de ordenadas (Y).

El gráfico lineal se diferencia del diagrama de barras y del histograma porque en estos últimos los valores de las variables que colocamos sobre el eje de abcisas (X) no deben empezar necesariamente por el valor cero ni éste ha de coincidir con el cero de ordenadas. Además, en ellos se utiliza el eje de ordenadas (Y) para

colocar las frecuencias de aparición de los valores de la variable, mientras que en el gráfico lineal esto no es así.

En la figura *2.15* se muestra un ejemplo de gráfico lineal.

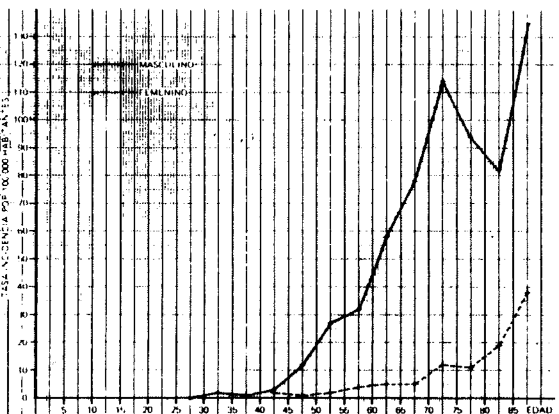


FIG. *2.15.— Tasas de incidencia del cáncer de vejiga según la edad y el sexo. Zaragoza, 1970-1974. Fuente:* GONZÁLEZ C.A.: *«Ocupación, tabaco, café y cáncer de vejiga en la comarca del Maresme». Servicio de Epidemiología y Estadística del Hospital de Sant Jaume i Santa Magdalena. Mataró, 1983, pág. 17.*

Gráfico semilogarítmico

La diferencia entre esta representación gráfica y la anterior radica en la escala utilizada en el eje de ordenadas (*Y*): en este caso, la logarítmica (por lo general, logaritmos decimales). En el eje de abcisas la escala empleada es lineal. De ahí que este tipo de representación se denomine «semilogarítmica».

Se utiliza cuando la variable que colocamos sobre el eje de ordenadas es susceptible de transformación logarítmica (véase capítulo *2.6*). Un ejemplo al respecto lo constituye la representación gráfica de las tasas de mortalidad por una enfermedad a lo largo de un determinado periodo de tiempo; en tal caso, la variable «tiempo» se colocará sobre el eje de abcisas y la variable «tasa de mortalidad por enfermedad» sobre el de ordenadas. En la figura *2.16* se ilustra este ejemplo.

El uso de escalas logarítmicas depende primordialmente de las variables en sí mismas y de los valores observados para cada una de ellas. Si se utilizan logaritmos decimales, como en el ejemplo anterior, los ciclos se repetirán cada 10 unidades.

Gráfico logarítmico

En el caso de los gráficos logarítmicos, ambas variables están representadas en escalas logarítmicas, y por lo general en logaritmos decimales. Un ejemplo de papel para gráficos logarítmicos puede verse en la figura *2.17*.

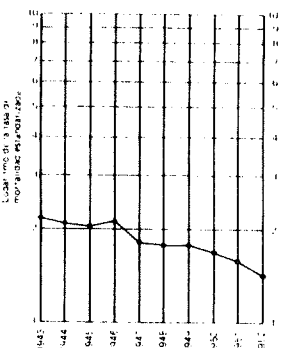


FIG. *2.16. Logaritmo de la tasa de mortalidad estandarizada por edad y sexo de las enfermedades infecciosas del tubo digestivo en España (1943-1952). Fuente: Movimiento Natural de la Población, 1943-1952. Instituto Nacional de Estadística, Madrid.*

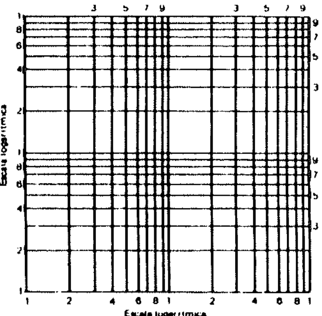


FIG. *2.17. — Papel para gráficos logarítmicos (dos ciclos en cada eje).*

V.— ERRORES EN LAS REPRESENTACIONES GRÁFICAS

Como en toda actividad humana, en la representación de la información mediante gráficos pueden cometerse errores que pueden ser debidos al diseño, construcción o interpretación de los mismos. El investigador debe procurar que la información que presenta mediante gráficos no sea errónea, confusa, ambigua o de pie a falsas interpretaciones.

Para evitar cometer errores debe tenerse en cuenta:

1.— El comienzo del eje de ordenadas siempre debe ser el 0.

Si las frecuencias se presentan a partir de un determinado valor de la escala, muy alejado del origen del eje, puede hacerse una muesca en el mismo («romper el eje») con objeto de utilizar una nueva escala más acorde con la situación.

Así, por ejemplo, si representamos los datos de la tabla *2.11* en un gráfico con el eje vertical completo obtendremos la figura *2.18*, mientras que si los representamos con el eje roto por una muesca, obtendremos la figura *2.19*.

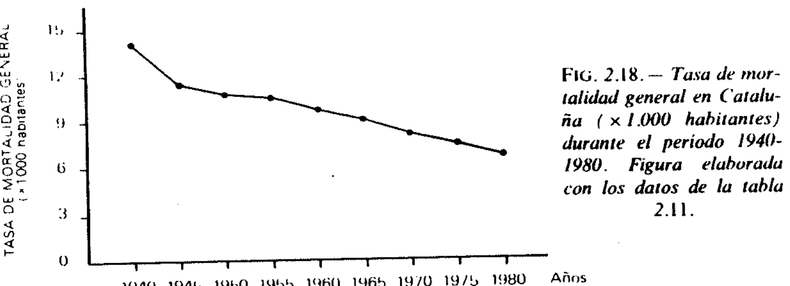


FIG. 2.18. — *Tasa de mortalidad general en Cataluña (× 1.000 habitantes) durante el periodo 1940-1980. Figura elaborada con los datos de la tabla 2.11.*

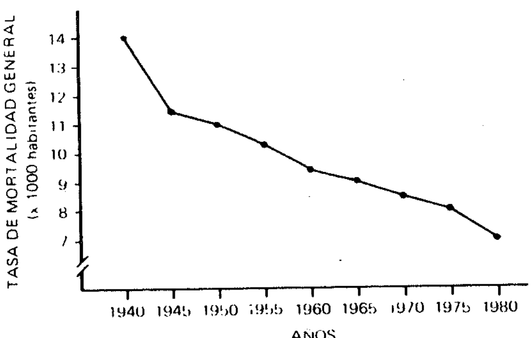


FIG. 2.19. — *Tasa de mortalidad general en Cataluña (× 1.000 habitantes) durante el periodo 1940-1980.*

Obsérvese como en el primer caso (figura 2.18) el descenso de la tasa de mortalidad parece mucho menos acusado que en el segundo (figura 2.19), siendo así que en ambos gráficos se representan los mismos datos originales.

Esto no constituye propiamente ningún error en la elaboración del gráfico puesto que en ambos casos se ha partido de 0 en el comienzo del eje de ordenadas. En todo caso puede inducir a error al lector poco avezado que no distingue las diferencias en la escala utilizada (¡Importancia de las escalas!.)

2.— En ocasiones se presentan datos continuos con un diagrama de barras. Esto constituye un grave error, por cuanto los diagramas de barras sólo sirven para las variables cualitativas o cuantitativas discretas, mientras que los histogramas son aplicables únicamente a las variables cuantitativas continuas.

3.— La construcción de los histogramas debe ser cuidadosa y detallada, recordando que son las *áreas* de cada rectángulo las que reflejan las frecuencias y no propiamente las alturas.

Un ejemplo al respecto se muestra en las figuras 2.20 y 2.21, construidas a partir de los datos de la tabla 2.14.

4. — Los pictogramas son representaciones gráficas que atraen en gran manera la visión del espectador. Por ello, si están mal diseñados darán una imagen falsa de la información que en ellos se ilustra. Es importante destacar que las áreas de los dibujos de dichos gráficos deben corresponderse con la magnitud real del fenómeno observado, sin hipertrofiarlo ni infravalorarlo.

TABLA 2.14. **Distribución por edades de los varones fallecidos a causa de leucemia**

EDAD (EN AÑOS)	EFECTIVOS OBSERVADOS (n_i)	LÍMITES DEL INTERVALO	PMI (x_i)	AMPLITUD DEL INTERVALO	EFECTIVOS POR AÑO DE EDAD
0	1	0-1	0,5	1	1,0
1-5	4	1-6	3,5	5	0,8
6-10	3	6-11	8,5	5	0,6
11-20	6	11-21	16	10	0,6
21-40	5	21-41	31	20	0,25
41-60	14	41-61	51	20	0,7
61-70	14	61-71	66	10	1,4
>70	19	71-100	85	29	0,66

Fuente: Registro de tumores del año 1978. Servicio de Oncología del Hospital de San Pablo, Barcelona. DOMÉNECH J. M.: «Descripción estadística de las variables». *MIR*, 1980, *2*, 534.

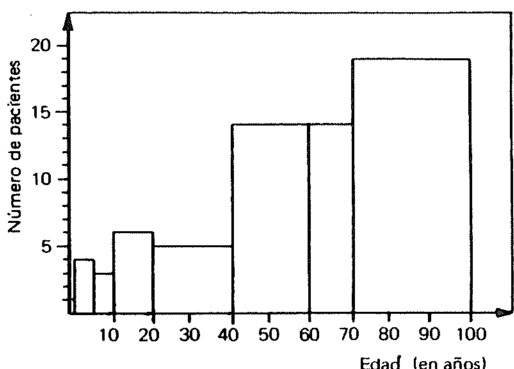


FIG. 2.20. — *Histograma incorrecto de los datos de la tabla 2.14.*

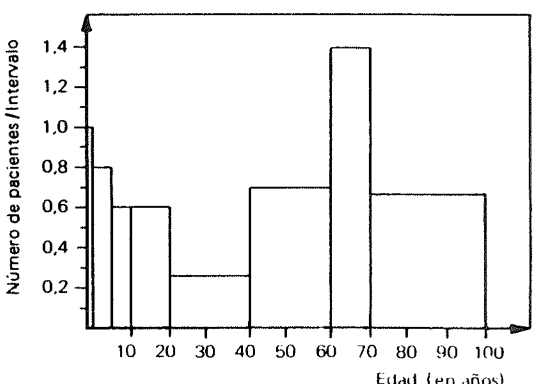


FIG. 2.21.— *Histograma correcto de los datos de la tabla 2.14.*

5.— Las escalas utilizadas en la presentación de la información recogida deben explicitarse al máximo. No es lo mismo presentar la evolución anual de las ventas de un determinado producto en kilogramos que en toneladas. La figura *2.21*, construida sobre los datos de la tabla *2.11*, parece indicar una disminución de la mortalidad mucho mayor que la reflejada en la figura *2.18*.

6.— Es imprescindible la rotulación de los gráficos. Cada representación gráfica debe tener un título o texto que explique su contenido. Además, si se emplean ejes de coordenadas, es necesario que estén rotulados y que en ellos se indiquen las unidades de medida utilizadas.

7.— Finalmente, no es pertinente ni adecuado abusar de la escala cromática ni añadir más información de la estrictamente relativa al gráfico empleado.

VI.— ESCALAS DE MEDIDA

Las escalas de medida se definen como un sistema para la asignación de números a los objetos o fenómenos que se estudian. En otras palabras, los fenómenos que se analizan deben poder manejarse como los números que los representan.

Teniendo en cuenta que cada subtipo asume todo lo dicho para el anterior, las escalas de medida se clasifican en (tabla *2.1*): o variables cualitativas

a) **Escala nominal.**— La operación matemática que realiza este sistema de medida es la *clasificación* de los individuos según particularidades de naturaleza cualitativa; determina únicamente la igualdad o desigualdad. Los atributos constituyen caracteres mensurables por esta escala. El sexo, la profesión y el estado civil son ejemplos concretos de atributos que se miden en escala nominal.

b) **Escala ordinal.**— La operación matemática que realiza este sistema de medida es la *ordenación* de los individuos; es decir, se establece entre los individuos estudiados una relación de mayor, igual o menor que (orden). Esta ordenación puede ser creciente o decreciente. Como ejemplos de variables que se miden en esta escala cabe citar el pronóstico vital, el tipo de bocio según el tamaño de la glándula tiroides, la prueba de Apgar.. (variables clasificables por rango)

c) **Escala de intervalo.**— La operación matemática que realiza este sistema de medida es la comprobación de la *igualdad o desigualdad de las distancias entre los valores que toman los individuos estudiados*. Esto supone la utilización de instrumentos de medida y la cuantificación propiamente dicha de los valores numéricos, que se podrán agrupar en «intervalos». Los caracteres medidos en esta escala deberán ser cuantitativos, pero las variables no tendrán un cero absoluto sino que éste será arbitrario o relativo. Así, la escala de intervalo asume la clasificación, ordenación y establecimiento de distancias conocidas entre los valores de la variable estudiada. La temperatura corporal en grados centígrados es un ejemplo de variable cuantitativa que se mide en esta escala. Así por ejemplo, un objeto a 40° C y otro a 39° C tienen la misma diferencia de «calor» que un objeto que esté a 29° C y otro a 28° C. En cambio, al no disponer de un cero absoluto, un objeto a 40° C no tiene el doble de «calor» que otro a 20° C.

d) **Escala de proporción o razón.**— La operación matemática que realiza este sistema de medida es el *establecimiento de proporciones o razones entre valores de la variable*, lo que supone un mayor grado de cuantificación de la variable y la existencia de un cero absoluto para ella. Esta escala asume todas las operaciones matemáticas de las anteriores: clasifica, ordena, establece intervalos y determina la igualdad de razones o proporciones entre los intervalos establecidos. Ejemplos de variables cuantitativas mensurables por escala de proporción son la talla de los individuos, la temperatura en grados absolutos o Kelvin y el nivel de glucosa plasmática. En el ejemplo de la temperatura, ahora sí se puede decir que un objeto a 400° K tiene el doble de «calor» que otro objeto a 200° K. De todas maneras, en las ciencias biológicas a menudo es difícil determinar la escala de medida (de intervalo o de razón o proporción) con que debe tratarse una variable cuantitativa.

TABLA *2.1.* — **Escalas de medida**

ESCALAS DE MEDIDA	OPERACIONES LÓGICAS	CONDICIONES	ESTADÍSTICOS APLICABLES
Nominal	Establecer la igualdad o desigualdad	Permutación	Frecuencias Moda
Ordinal	Determinar lo mayor, lo igual o lo menor	Mantenimiento del orden	Medianas Percentiles Correlación ordinal
De intervalo	Determinar la igualdad o diferencia de intervalo	Unidad constante	Media aritmética Desviación típica Correlaciones
De razón o proporción	Establecer igualdad de razones	Cero absoluto	Media geométrica Coeficiente de variación

Fuente: STEVENS S. S.: «On the theory of scales of measurement». *Science*, 1946, *103*, 677-680 (modificada).